

SEQUENCE LISTING

(1) GENERAL INFORMATION:



- (i) APPLICANT: DECKER, Heinrich
- (ii) TITLE OF INVENTION: ISOLATION OF THE BIOSYNTHESIS GENES FOR PSEUDO-OLIGOSACCHARIDES FROM STREPTOMYCES GLAUDESCENS GLA.O AND THEIR USE
- (iii) NUMBER OF SEQUENCES: 13
- (iv) CORRESPONDENCE ADDRESS:
  - (A) ADDRESSEE: FOLEY & LARDNER
  - (B) STREET: 3000 K Street, N.W.
  - (C) CITY: Washington
  - (D) STATE: D.C.
  - (E) COUNTRY: U.S.A.
  - (F) ZIP: 20007-5109
- (v) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Floppy disk
  - (B) COMPUTER: IBM PC compatible
  - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
  - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 09/194,905
  - (B) FILING DATE: 19-JUL-1998
  - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: WO PCT/EP97/02826
  - (B) FILING DATE: 30-MAY-1997
- (viii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: DE 19622783.6
  - (B) FILING DATE: 07-JUN-1996
- (ix) ATTORNEY/AGENT INFORMATION:
  - (A) NAME: Granados, Patricia D.
  - (B) REGISTRATION NUMBER: 33,683
  - (C) REFERENCE/DOCKET NUMBER: 026083/0193
- (ix) TELECOMMUNICATION INFORMATION:
  - (A) TELEPHONE: (202) 672-5300
  - (B) TELEFAX: (202) 672-5399

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 22 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: other nucleic acid
  - (A) DESCRIPTION: /desc = "Primer"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CSGGSGSSGC SGGSTTCATS GG

22

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "Primer"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GGGWVCTGGY VSGGSCCGTA GTTG

24

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 546 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CCCGGGGGGGG GCGGGGGTTCA TCGGCTCCGC CTACGTCCGC CGGCTCCTGT CGCCCGGGGC	60
CCCCGGCGGGC GTCGCGGTGA CCGTCTCGA CAAACTCACC TACGCGGGCA GCTCGCCCG	120
CCTGCACGGG GTGCGTGACC ATCCCGGCT CACCTTGTG CAGGGCGACG TGTGGACAC	180
CGCGCTCGTC GACACGCTGG CCGCGGGCA CGACGATC GTGCACTTG CGGCGAGTC	240
GCACGTGAC CGCTCCATCA CCGACAGGG TGCTTCACC CGCACCAAG TGTGSGGAC	300
CCAGGTCTTG CTGACGCGG CGCTCGGCA CGGTGTGGC ACCCTGTGC AGTCTCCAC	360
CGACGAGGTG TACGGCTCCC TCCCGCAGG GCGCGCGGG GAGAGCGACC CCTGCTCCC	420
GACCTCGCGG TACGGGGGT GGAAGGGGG CTGGGACCTC ATGGGCTTG CCCAGACCG	480
CACCCACGGC CTGGACGTCC GGGTGACCG CTGTTGGAAC AACTACGGG CGCACCATT	540
CCCGGG	546

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 541 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
CCCCGGGTGC TGGTAGGGGC CGTAGTTGTT GGAGCAGCGG GTGATGCGCA CGTCCAGGCC      60
GTGGCTGACG TGCATGGCCA GCGCGAGCAG GTCGCCCCGAC GCCTTGGAGG TGGCATAGGG      120
GCTGTTGGGG CGCAGCGGCT CGTCCTCCGT CCACGACCCC GTCTCCAGCG AGCCGTAGAC      180
CTCGTCGGTG GACACCTGCA CGAAGGGGGC CACGCCGTGC CGCAGGGCCG CGTCGAGGAG      240
TGTCTGCGTG CCGCCGGCGT TGGTCCGCAC GAACGCGGCG GCATCGAGCA GCGAGCGGTC      300
CACGTGCGAC TGGGCGGCGA GGTGCACGAC CTGCTCTTGG CCGGGCATGA CCGGTCGAC      360
CAGGTCCGCG TCGCAGATGT CGCCGTGGAC GAAGCGCAGC CGGGGGTGST CGCGGACCGG      420
GTCGAGGTTG GCGAGGTTGC CGGCGTAGCT CAGGGCGTCG AGCACGGTGA CGACGGCGTC      480
GGGCGGGCCG TCCGGACCGA GGAGGGTGCG GACGTAGTGC GAGCCCATGA ACCCGGCCGC      540
C
```

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 180 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
Ala Ala Gly Phe Met Gly Ser His Tyr Val Arg Thr Leu Leu Gly Pro
1           5           10           15
Asp Gly Pro Pro Asp Ala Val Val Thr Val Leu Asp Ala Leu Ser Tyr
20           25           30
Ala Gly Asn Leu Ala Asn Leu Asp Pro Val Arg Asp His Pro Arg Leu
35           40           45
Arg Phe Val His Gly Asp Ile Cys Asp Ala Asp Leu Val Asp Arg Val
50           55           60
Met Ala Gly Gln Asp Gln Val Val His Leu Ala Ala Glu Ser His Val
65           70           75           80
Asp Arg Ser Leu Leu Asp Ala Ala Ala Phe Val Arg Thr Asn Ala Gly
85           90           95
Gly Thr Gln Thr Leu Leu Asp Ala Ala Leu Arg His Gly Val Ala Pro
100          105          110
```

Phe	Val	Gln	Val	Ser	Thr	Asp	Glu	Val	Tyr	Gly	Ser	Leu	Glu	Thr	Gly
	115						120					125			
Ser	Trp	Thr	Glu	Asp	Glu	Pro	Leu	Arg	Pro	Asn	Ser	Pro	Tyr	Ala	Thr
	130					135					140				
Ser	Lys	Ala	Ser	Gly	Asp	Leu	Leu	Ala	Leu	Ala	Met	His	Val	Ser	His
145					150					155					160
Gly	Leu	Asp	Val	Arg	Ile	Thr	Arg	Cys	Ser	Asn	Asn	Tyr	Gly	Pro	Tyr
			165						170					175	
Gln	His	Pro	Gly												
			180												

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 181 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Pro	Gly	Gly	Ala	Gly	Phe	Ile	Gly	Ser	Ala	Tyr	Val	Arg	Arg	Leu	Leu
1			5					10						15	
Ser	Pro	Gly	Ala	Pro	Gly	Gly	Val	Ala	Val	Thr	Val	Leu	Asp	Lys	Leu
		20				25						30			
Thr	Tyr	Ala	Gly	Ser	Leu	Ala	Arg	Leu	His	Ala	Val	Arg	Asp	His	Pro
	35					40						45			
Gly	Leu	Thr	Phe	Val	Gln	Gly	Asp	Val	Cys	Asp	Thr	Ala	Leu	Val	Asp
50					55					60					
Thr	Leu	Ala	Ala	Arg	His	Asp	Asp	Ile	Val	His	Phe	Ala	Ala	Glu	Ser
65				70					75					80	
His	Val	Asp	Arg	Ser	Ile	Thr	Asp	Ser	Gly	Ala	Phe	Thr	Arg	Thr	Asn
			85						90					95	
Val	Leu	Gly	Thr	Gln	Val	Leu	Leu	Asp	Ala	Ala	Leu	Arg	His	Gly	Val
		100					105					110			
Arg	Thr	Leu	Val	His	Val	Ser	Thr	Asp	Glu	Val	Tyr	Gly	Ser	Leu	Pro
	115						120					125			
His	Gly	Ala	Ala	Ala	Glu	Ser	Asp	Pro	Leu	Leu	Pro	Thr	Ser	Pro	Tyr
	130					135					140				
Ala	Ala	Ser	Lys	Ala	Ala	Ser	Asp	Leu	Met	Ala	Leu	Ala	His	His	Arg
145					150				155						160
Thr	His	Gly	Leu	Asp	Val	Arg	Val	Thr	Arg	Cys	Ser	Asn	Asn	Tyr	Gly
			165					170						175	
Pro	His	Gln	Phe	Pro											
			180												

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6854 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CTGCAGGGTT CCTGGTGCA CGACCCGCC CTGGTCGACG ACCAGGGCGC TGTGCGAGAT	60
CGCGCGGATG TCGGCGATGT GCTGGCTGCT CAGCAGCAGC CTGCTGCCCA GTTCCGGTG	120
GGGCGGGTTG ACCAGCCGGC GCACCGCGTC CTTCAGCACC ATGTGAGAGC CGATCGTGGG	180
CTCGTCCCAg AACAGCACGG CCGGGTCTG CAGCAGGCTC GCCGCGATCT CGGCGCGCAT	240
GGGCTGTCCG AGGCTGAGCT GCCGCACGGG GGTGGACCCC AGCGCGTCGA TGTGAGGAG	300
GTCCCGGAAC ATGGCGAGGT TGGCCGGTA GACCGGTCCG GGGATGTCGT AGATGCGGCG	360
CAGSATGCGG AAGGAGTCGG GTACGACAG GTCCACCAg AGCTGGCTGC GCTGGCCGAA	420
GACGACGCG ATCTGTGCGG CGTTGCGCTG CCGGTGCCGG TAGGGCTCCA GCCCGGCGAC	480
CGTGCAGCGG CCGGAGGTGG GGGTCATGAT GCGGTCAGC ATCTTGATCG TGGTCGACTT	540
GCCGGCTCCG TTGGCGCCGA TGTAGGCGGT CTTCGTGCGG GCGGTATCT CGAAGGAGAC	600
GTGTCGAGC GGGCGCAGGA CGGGTACCG GGGGTTCAGG AGGGTGGAGA GGTGCCGAG	660
CAGGCCGGGC TGCGTTCCG CCAGCCGGAA CTCCTTGACG AGGTGTTCGG CCACGATCAC	720
GCGATCACCC GTTCGACGGC CGTCTCCAGC AGGCGCAGGC CCTCGTCGAG CAGCGCCTCG	780
TGGAGGGTGA ACGGCGGTGC CAGCCGCAGG ATGTGGCCGC CCAGGGAGGT GCGCAGCCCC	840
AGGTGAGGG CGGTGGTGTA GACGGCCCGG GCGGTCTCGG GGGCGGGTGC CCGGCCGACG	900
GCGTCGGTGA CGAACTCCAG GCGCCACAGC AGTCCGAGGC CGGTACCTG GCGAGCTGG	960
GGGAAGCGGG ATCCAGGGC GCGCAGCCGC TCCTGGATGA GTTCGCGAG GACGCGCAGG	1020
CGGTGATCA GCGGTCCGG CTGACGACC TCCAGCGTGG CGCGGGGGG GGGATCCCC	1080
AGTGGGTTGC TCGGTACGT CGAGGCGTAC GCGCCGGGGT GGCGGCTCC GGCCTGCGCA	1140
GCTTCGCGC GTGGGGCCAG CAGGGGGAAG GGAATCCGC TCGGGTGCC CTTCGACAGC	1200
ATGCCCAGGT CGGGTCGAT GCGAACAAT TGGCTGGCGA GGAAGGCGCC GGTGCGCCCG	1260
CGCGGGTGA GAACTGCTC GCGGACGAG AGCAGCGCCG CGTCCCGGCA GCGCGCGGG	1320
ATCCGCTCC AGTAGCGGG GGGCGGACG ATGACGCTG CCGCGCCGAG GACGGGTTCS	1380
AAGACCAGGG CCGAGACGTT GGGCTTCTCC GCGATGTGCC GCGGCACGAG GGTGCGCAC	1440

CGCACGTGCG	ACGAGGGGTA	CTCCAGGCCC	AGGGGACAGC	GGTAGCCAGT	AGGGGCTGTA	1500
GCCAGCACGC	TGTTGCCGCT	GAAGGCCTGG	TGGCCGATGT	CCCACTGGAC	CAGCATCCGG	1560
GCGCCCATGG	TCTTGCCGTG	GAAGCCGTGG	CGCAGGGCGC	AGATCCGGTT	GCGGCCCGGC	1620
GCGGCGGTGG	CCTGGACGAC	CCGCAGGGCG	GCCTCGACCA	CCTCCGCGCC	GGTGGAGAAG	1680
AAGGCGTAGG	TGTCGAGCTG	TTGGGGCAGC	AGCCTGGCGA	GCAGTTCCAG	CAGGCCGGGG	1740
CGGTCCGGCG	TGGCGCTGTC	GTGGACGTTG	CACAGGCGGC	GGGCCTGGGT	GGTGAGTGCC	1800
TCGACGACCT	CCGGGTGCCC	GTGGCCCACT	GACTGGGTGA	GGGTCCCGGC	CGCGAAGTCG	1860
AGGTACTGGT	TGCCGTCCAG	GTGGGTGAGA	ACGGGACCGC	GTCCCTCGGC	GAAGAACCAG	1920
CGTCCGTGGA	CGGCTTCCTC	GGAGGCGCCC	GGCGCCAGGT	GGCGGGGCTC	CCGTGCCAGG	1980
TGCTGTGTCT	GGGTAAAGCC	TGTCATCGCT	GGCTCTGCTC	GTGGGACCGG	CTGACGGGAT	2040
CGCCGGCGAA	CTGCGTTGTC	GCGCACCACG	GTTGGGGCGG	CTCGGGGCTG	AGTCAAAACAC	2100
TTGAACACAC	ACCGCTGCAA	GAGTTTGCGG	GTTGTTTCAG	AAAGTTGTTG	CGAGCGGGCC	2160
CGGCACTCTG	GTTGAGTCGA	CGTGCTTACG	GCGCCACCCAC	GGCTCACGTT	CGAGGAGGGA	2220
CCTGTGAGAA	CAAGCCCGCA	GACCGACCGG	CTCCCGGGGA	GGCCGAGGTG	AAGGCCTGCG	2280
TCCTGGCAGG	TGGAACCGGC	AGCAGACTGA	GGCGGTTCCAC	CCACACCGCC	GCCAAGCAGC	2340
TGCTCCCAT	CGCCAACAAG	CCCGTGCTCT	TCTACGGGCT	GGAGTCCCTC	GCGCGGGCGG	2400
GTGTCCGGGA	GGCCGGGGTC	GTGGTGGGCG	GCTACGGGCG	GGAGATCCGC	GAAGTCACCG	2460
GCGACGGCAC	CGCGTTCCGG	TTACGCATCA	CCTACCTCCA	CGAGCCCGCG	CCGCTCGGTC	2520
TGGGCGACGC	GSTGGCGATC	GCCCCGGGCT	TGCTGGGGGA	CGACGACTTC	CTGCTGTACC	2580
TGGGGGACAA	CTACCTGCCC	CAGGCGGTCA	CCGACTTGCG	CCGCCAATCG	GCCGCGGATC	2640
CGCGGGCGGC	CCGGCTGCTG	CTCACCCCGG	TGGGGGACCC	GTCCGCTTTC	GGGCTCGCGG	2700
AGGTGACGCG	GGACGGGAAC	GTGCTGGGCT	TGGAGGAGAA	ACCCGACGTC	CGCGGCAGCT	2760
CGCTCGCGCT	CATCGGCGTG	TACGCCTTCA	GCCCCGGCGT	CCACGAGGCG	GTACGGGGCA	2820
TCACCCCTTC	CGCCCGCGGC	GAGCTGGAGA	TCACCCACGC	CGTGCACTGG	ATGATCGAGC	2880
GGGGCTGCG	CGTACGGGCC	GAGACCACCA	CCCGGGCCTG	GGCGGACACC	GGCAGCGCGG	2940
AGGACATGCT	GGAGGTCAAC	CGTCACGTCC	TGGACGGACT	GGAGGGCCGC	ATCGAGGGGA	3000
AGGTGACGCG	GCACAGCACG	CTGGTCGGCC	GGGTCCGGGT	GGCCGAAGGG	GCGATCGTGC	3060
GGGGGTGACA	CGTGGTGGGC	CCGGTGGTGA	TGGGCGCGGG	TGCCGTGCTC	AGCAACTGCA	3120
GTGTGGGCCC	GTACACCTCC	ATCGGGGAGG	AATGCGGGGT	CGAGGACAGC	GCCATCGAGT	3180
ACTCGGTCTT	GCTGGCGGGC	GCCCCAGGTG	AGGGGGGCTC	CCGCATCGAG	GGGTGCTTCA	3240
TGGGCGGCGG	CGCGGTGGTC	GGGCGGGGCC	CCGCTGTCCC	GCAGGCTCAC	CGACTGCTGA	3300
TGGGCGACCA	CAGCAAGGTG	TATCTCACCC	CATGACCACG	ACCATCCTCG	TCACCGGGGG	3360

AGCGGGCTTC ATTCGCTCCG CCTACGTCCG CCGGCTCCTG TCGCCCGGGG CCCCCGGCGG	3420
CGTCGCGGTG ACGTCCCTCG ACAAACCTAC CTACGCCGGC AGCTCGGCC GCTGCAACGC	3480
GCTGCGTGAC CATCCCGGGC TCACCTTCGT CCAGGGCGAC GTGTGCGACA CCGGCTCGT	3540
CGACACGCTG GCGCGCGGGC ACGACGACAT CGTGCACTTC GCGGCGAGT CGCACGTGCA	3600
CCGCTCCATC ACGACAGCG GTGCCTTCAC CCGCACCAAC GTGCTGGGCA CCCAGGTCT	3660
GCTCGACGGC GCGCTCCGGC AGCTGTGCG CACCTTCGTG CAGGTCTCCA CGACGAGGT	3720
GTACGGCTCC CTCGCCACG GGGCGCGCGC GGAGAGCGAC CCGTGTCTTC CGACCTCGGC	3780
GTACGCGGGG TCGAAGGGG CCTCGGACT CATGGCGTC GCGCACACC GCACCCACGG	3840
CCTGGACGTC CCGGTGACCC GGTGTTCGAA CAACTTGGG CCCCACAGC ATCCCGAGAA	3900
GCTCATACCG CGCTTCCTGA CCAGCTCTCT GTCCGGGGG ACCGTTCGCC TCTACGGCGA	3960
CGGGCGGCAC GTGCGGACT GGCTGCACGT CGACGACAC GTGAGGGCGG TCGAACTCGT	4020
CCGCGTGTG GCGCGCGCGG GAGAGATCTA CAACTCGGG GCGGGACCT CGCTGCCCAA	4080
CCTGGAGCTC ACGCAACGGT TGCTCGACT GTGCGGGCG GCGCGGAGC GCATGCTCCA	4140
CGTCGAGAAC CGCAAGGGG ACGACCGGG CTACGGGTC GACCAACCA ASATCACCGC	4200
GGAACTCGGT TACCGGCCGC GCACCGACTT CCGGACCGG CTGGCGACA CCGCGAAGTG	4260
GTACGAGCGG CACGAGGACT GGTGGCGTC CCTGCTGCC GCGACATGAC GTGCGGCGG	4320
ACCGCAACCA CCGGCCCCGG CCGGCACAC GCGCGCGGG GCGGTGGCC GCGCGGTGAG	4380
CGTCCGTGAG CCGGGCGCGG GCGGCCCGC GCGCGCGGG CCGTGGACC CCGGACGAC	4440
AGTTCGGCA TGAAGACGAA TTCGGTGGC GCGGGCGGG TTGCGCTCAT CTCTCGAGC	4500
AGTGGCTCCA CCGGACCTG CCGATGCGC TTGAGGGGT GTGTGATGT GTTAGGGGA	4560
GGGTCCGTGA AGGCTATGAG CCGCGAGTGG TCGAAGCGA CCAGCGAGAT GTTACCGGA	4620
ACCGTGAGAC CCGGCGGGG CCGGGCGCG ACGGGGCGA GCGCATCAT GTGCTGGG	4680
CACATGACCG CGGTGCAACC CAGGTGATC AGCGGGAGG CCGGCGCTG GCGCGCTCC	4740
AGGAGAGACA GCGAGTGTG CACGAGCTC TCGGATCCC GCGCGACAC TCCAGGTGC	4800
TCCCGCAGGC CCGGCGGAA CCGCTGATC TTGCTGCA CCGGACGAA GCGGGCGGGC	4860
CGGACGGGA GCGGACGGC CTGCTGCCG AGCTCGCCA GGTGCGCAC GCGGAGGCG	4920
ATCGCGGCC GGTCTCGGG GGAGACGAA GGTGCTCGA TCGGGGCGA GAACTGCTC	4980
ACGAGGACGA AGGGACCTG CCGCTGCTC AGCGGGCGT ACGTCCGGT CTCGGGGTG	5040
GTGTCCGGT GCAGTCGGA GACGAAGATG ATGCGGACA CCGGCGGTC CACGAGCATC	5100
TCCGTGASTT CTTCTCGGT CGAGCGGCC GGGGTCTGG TGGGAGCAC GGGGTGTAG	5160
CCCTGACGG TGAGCGCTG CCGATCACG TGGGACGTG GCGGGAAGAA GGGGTGTGC	5220
AGTTCGGGG TGACAGTCC GACGAGCTG GCGGGGGT GTGGGCGGG GTGCTGTAG	5280

CCCAGCGCGT CCAGTGCGGT CAGCACCGAG TCGCGGGTGC CGGTGGCCAC ACCGCGCGCA	5340
CCGTTACGCA CCGGGCTGAC CGTGGCCTTG CTGACGCCCG CCGGGGCTGC GATGTGGGG	5400
AGCCGATGG TCATGGCAAC GCACTCTACC TGTGGGGGCG TCAGGGGCTG CCCACCGGCG	5460
GCGGAACCGG CGGACTGGG GGCACGGCCC GTCCGCCGCC CACGGACCAC GCGCCCGAAA	5520
CGATGGCTGA AAATGCTTGC AGCAAATTGC CGCAACGTCT TTCGGCGGCT TTTCGATCCT	5580
GTTAGTTTC TGGCAACCCC GCGGCCGCGC AGAAGCGGTT GCGGTGAGGC GTCCAGACCT	5640
CCGCCCGATT CCGGGATCAC TCAGGGGAGT TCACAATGG GCGTGGCATT GCGGCCACCG	5700
CGGTGTTCCG GGTGTGGCC ATGACGGCAT CGGCGTGTGG CCGGGGCGAC AACGGCGGAA	5760
GCGGTACCGA CGGGGGCGGC ACGSAGCTGT CCGGGACCGT CACCTTCTGG GACACGTCCA	5820
ACGAAGCGGA GAAGGGGAGG TACCAGGCC TCGGGGAGGG CTTCGAGAAG GAGCACCGGA	5880
AGGTGGACGT CAAGTACGTC AACGTCCCGT TCGGGGAGGC GAACGCCAAG TTCAAGAAAG	5940
CCGCGGGGGG CAATCCGGT GCGCCGAGG TGATGGTAC GGAGGTGCGC TGGGTGCGGG	6000
ACTTGCGCAG CATCGGCTAC CTCGCCCGCG TCGACGGCAC GCGCGCCCTC GACGACGGGT	6060
CGSACCACT TCCCGAGGCG GGCAGACCA GSTACGAGGG GAAGACCTAC GCGGTCCCGC	6120
AGSTGATGGA GATCTGGGG CTCTTCTACA ACAAGGAACT GCTGAGGAAG GCGGTGTGG	6180
AGGTGCCGGG CTCCTCGCC GAGCTGAAGA CGGCCGCCCG CGAGATCACC GAGAAGACCG	6240
GCGCGAGCGG CCTCTACTGC GGGCGACGA CCGTACTTG GTTCTGCCC TACCTCTAGG	6300
GGGAGGGGGG CGAAGTGGTC GACGAGAAAG ACAAGACCGT CACGGTCGAC GACGAAGCGG	6360
GTGTGCGCGC CTACCGGCTC ATCAAGGACC TCGTGGACAG CAAGGCGGCC ATCACCGAGG	6420
CGTCGAGGG CTGGAAACAAC ATGCAGAAG CCTTCAAGTC GGGCAAGGTC GCCATGATGG	6480
TCAACGGCCC CTGGGCCATC GAGGACGTCA AGGCGGGAGC CCGCTTCAAG GACGCGGGCA	6540
ACCTGGGGGT CGGCCCGCTC CCGGCCGGCA GTGCCGGACA GGGCTCTCCC CAGGGCGGGT	6600
GGAACCTCTC GGTGTACGG GGCTCGAAGA ACCTCGACGC CTCCTACGCC TTCGTGAAGT	6660
ACATGAGCTC CGCAAGGTTG CAGCAGCAGA CCACCGAGAA GCTGAGCCTG CTGCCACCC	6720
GCACGTCCGT CTACGAGGTC CCGTCCGTGG CGGACAACGA GATGGTGAAG TTCTTCAAGT	6780
CGGCCGTGGA CAAGGCCCTC GAACGGCCGT GGATCGCCGA GGGCAATGCC CTCTTCGAGG	6840
CGATCGGGCT GCAG	6854

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 240 amino acids
- (B) TYPE: amino acid
- (C) SIMILARITY:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein



(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Val Ile Val Ala Glu His Leu Val Lys Glu Phe Arg Leu Ala Glu Arg  
1 5 10 15  
Glu Pro Gly Leu Leu Gly Ser Leu Ser Thr Leu Leu Thr Arg Arg Tyr  
20 25 30  
Arg Val Val Arg Ala Val Asp Asp Val Ser Phe Glu Ile Pro Ala Gly  
35 40 45  
Thr Lys Thr Ala Tyr Ile Gly Ala Asn Gly Ala Gly Lys Ser Thr Thr  
50 55 60  
Ile Lys Met Leu Thr Gly Ile Met Thr Pro Thr Ser Gly Arg Cys Thr  
65 70 75 80  
Val Ala Gly Leu Glu Pro Tyr Arg His Arg Gln Arg Asn Ala Arg Thr  
85 90 95  
Ile Gly Val Val Phe Gly Gln Arg Ser Gln Leu Trp Trp Asp Leu Ser  
100 105 110  
Val Pro Asp Ser Phe Arg Ile Leu Arg Arg Ile Tyr Asp Ile Pro Gly  
115 120 125  
Pro Val Tyr Arg Arg Asn Leu Ala Leu Phe Arg Asp Leu Leu Asp Ile  
130 135 140  
Asp Ala Leu Gly Ser Thr Pro Val Arg Gln Leu Ser Leu Gly Gln Arg  
145 150 155 160  
Met Arg Ala Glu Ile Ala Ala Ser Leu Leu His Asp Pro Ala Val Leu  
165 170 175  
Phe Trp Asp Glu Pro Thr Ile Gly Leu Asp Met Val Leu Lys Asp Ala  
180 185 190  
Val Arg Arg Leu Val Asn Arg Ala His Arg Glu Leu Gly Thr Thr Val  
195 200 205  
Val Leu Thr Ser His Asp Ile Ala Asp Ile Ala Ala Ile Cys Asp Ser  
210 215 220  
Ala Leu Val Val Asp Gln Gly Arg Val Val His Gln Gly Thr Leu Gln  
225 230 235 240

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 429 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Met Thr Gly Leu Arg Gln Thr Gln His Leu Ala Arg Glu Ala Arg His  
1 5 10 15  
Leu Ala Pro Gly Ala Ser Glu Glu Ala Val His Gly Arg Arg Val Phe  
20 25 30  
Ala Glu Gly Arg Gly Pro Val Leu Thr Asp Leu Asp Gly Asn Gln Tyr  
35 40 45  
Leu Asp Phe Ala Ala Gly Thr Leu Thr Gln Ser Leu Gly His Gly His  
50 55 60  
Pro Glu Val Val Glu Ala Leu Thr Thr Gln Ala Arg Arg Leu Trp Asn  
65 70 75 80  
Val His Asp Ser Ala Thr Pro Asp Arg Ala Gly Leu Leu Glu Leu Leu  
85 90 95  
Ala Arg Leu Leu Pro Glu Gln Leu Asp Thr Tyr Ala Phe Phe Ser Thr  
100 105 110  
Gly Ala Glu Val Val Glu Ala Ala Leu Arg Val Val Gln Ala Thr Ala  
115 120 125  
Ala Pro Gly Arg Asn Arg Ile Cys Ala Leu Arg His Gly Phe His Gly  
130 135 140  
Lys Thr Met Gly Ala Arg Met Leu Val His Trp Asp Ile Gly His Gln  
145 150 155 160  
Ala Phe Ser Gly Asn Ser Val Leu Ala Thr Ala Pro Thr Gly Tyr Arg  
165 170 175  
Cys Pro Leu Gly Leu Glu Tyr Pro Ser Cys Asp Val Arg Cys Ala Thr  
180 185 190  
Leu Val Arg Arg His Ile Ala Glu Lys Pro Asn Val Ser Ala Leu Val  
195 200 205  
Phe Glu Pro Val Leu Gly Ala Ala Gly Val Ile Val Pro Pro Pro Gly  
210 215 220  
Tyr Trp Glu Arg Ile Ala Gly Ala Cys Arg Asp Gly Gly Val Leu Leu  
225 230 235 240  
Val Ala Asp Glu Val Leu Thr Gly Gly Gly Arg Thr Gly Ala Phe Leu  
245 250 255  
Ala Ser Glu Leu Phe Gly Ile Glu Pro Asp Leu Ala Met Leu Ser Lys  
260 265 270  
Gly Thr Ala Ser Gly Phe Pro Phe Ala Val Leu Ala Gly Arg Ala Glu  
275 280 285  
Ala Ala Gln Ala Gly Gly Gly His Pro Gly Ala Tyr Ala Ser Thr Tyr  
290 295 300  
Ala Ser Asn Pro Leu Gly Ile Ala Ala Ala Arg Ala Thr Leu Glu Val  
305 310 315 320  
Val Glu Arg Asp Arg Leu Ile Asp Arg Val Arg Val Leu Gly Glu Leu  
325 330 335

Ile Gln Glu Arg Leu Arg Ala Leu Glu Ser Arg Phe Pro Gln Leu Gly  
 340 345 350  
 Gln Val Arg Gly Leu Gly Leu Leu Trp Gly Leu Glu Phe Val Thr Asp  
 355 360 365  
 Ala Val Gly Arg Ala Pro Ala Pro Glu Thr Ala Arg Ala Val Tyr Thr  
 370 375 380  
 Thr Ala Leu Asp Leu Gly Leu Arg Thr Ser Leu Gly Gly His Ile Leu  
 385 390 395 400  
 Arg Leu Ala Pro Pro Phe Thr Leu Asp Glu Ala Leu Leu Asp Glu Gly  
 405 410 415  
 Leu Arg Leu Leu Glu Thr Ala Val Glu Arg Val Ile Ala  
 420 425

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 355 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS:
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Val Lys Ala Leu Val Leu Ala Gly Gly Thr Gly Ser Arg Leu Arg Pro  
 1 5 10 15  
 Phe Thr His Thr Ala Ala Lys Gln Leu Leu Pro Ile Ala Asn Lys Pro  
 20 25 30  
 Val Leu Phe Tyr Ala Leu Glu Ser Leu Ala Ala Ala Gly Val Arg Glu  
 35 40 45  
 Ala Gly Val Val Val Gly Ala Tyr Gly Arg Glu Ile Arg Glu Leu Thr  
 50 55 60  
 Gly Asp Gly Thr Ala Phe Gly Leu Arg Ile Thr Tyr Leu His Gln Pro  
 65 70 75 80  
 Arg Pro Leu Gly Leu Ala His Ala Val Arg Ile Ala Arg Gly Phe Leu  
 85 90 95  
 Gly Asp Asp Asp Phe Leu Leu Tyr Leu Gly Asp Asn Tyr Leu Pro Gln  
 100 105 110  
 Gly Val Thr Asp Phe Ala Arg Gln Ser Ala Ala Asp Pro Ala Ala Ala  
 115 120 125  
 Arg Leu Leu Leu Thr Pro Val Ala Asp Pro Ser Ala Phe Gly Val Ala  
 130 135 140  
 Glu Val Asp Ala Asp Gly Asn Val Leu Arg Leu Glu Glu Lys Pro Asp  
 145 150 155 160

Val	Pro	Arg	Ser	Ser	Leu	Ala	Leu	Ile	Gly	Val	Tyr	Ala	Phe	Ser	Pro
				165					170					175	
Ala	Val	His	Glu	Ala	Val	Arg	Ala	Ile	Thr	Pro	Ser	Ala	Arg	Gly	Glu
			180					185					190		
Leu	Glu	Ile	Thr	His	Ala	Val	Gln	Trp	Met	Ile	Asp	Arg	Gly	Leu	Arg
		195					200					205			
Val	Arg	Ala	Glu	Thr	Thr	Thr	Arg	Pro	Trp	Arg	Asp	Thr	Gly	Ser	Ala
		210					215				220				
Glu	Asp	Met	Leu	Glu	Val	Asn	Arg	His	Val	Leu	Asp	Gly	Leu	Glu	Gly
225					230					235					240
Arg	Ile	Glu	Gly	Lys	Val	Asp	Ala	His	Ser	Thr	Leu	Val	Gly	Arg	Val
				245					250					255	
Arg	Val	Ala	Glu	Gly	Ala	Ile	Val	Arg	Gly	Ser	His	Val	Val	Gly	Pro
			260					265					270		
Val	Val	Ile	Gly	Ala	Gly	Ala	Val	Val	Ser	Asn	Ser	Ser	Val	Gly	Pro
		275					280					285			
Tyr	Thr	Ser	Ile	Gly	Glu	Asp	Cys	Arg	Val	Glu	Asp	Ser	Ala	Ile	Glu
	290					295					300				
Tyr	Ser	Val	Leu	Leu	Arg	Gly	Ala	Gln	Val	Glu	Gly	Ala	Ser	Arg	Ile
305					310					315					320
Glu	Ala	Ser	Leu	Ile	Gly	Arg	Gly	Ala	Val	Val	Gly	Pro	Ala	Pro	Arg
				325					330					335	
Leu	Pro	Gln	Ala	His	Arg	Leu	Val	Ile	Gly	Asp	His	Ser	Lys	Val	Tyr
			340					345					350		
Leu	Thr	Pro													
			355												

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 325 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Met	Thr	Thr	Thr	Ile	Leu	Val	Thr	Gly	Gly	Ala	Gly	Phe	Ile	Arg	Ser
1				5				10					15		
Ala	Tyr	Val	Arg	Arg	Leu	Leu	Ser	Pro	Gly	Ala	Pro	Gly	Gly	Val	Ala
			20					25					30		
Val	Thr	Val	Leu	Asp	Lys	Leu	Thr	Tyr	Ala	Gly	Ser	Leu	Ala	Arg	Leu
			35				40					45			

His Ala Val Arg Asp His Pro Gly Leu Thr Phe Val Gln Gly Asp Val  
 50 55 60  
 Cys Asp Thr Ala Leu Val Asp Thr Leu Ala Ala Arg His Asp Asp Ile  
 65 70 75 80  
 Val His Phe Ala Ala Glu Ser His Val Asp Arg Ser Ile Thr Asp Ser  
 85 90 95  
 Gly Ala Phe Thr Arg Thr Asn Val Leu Gly Thr Gln Val Leu Leu Asp  
 100 105 110  
 Ala Ala Leu Arg His Gly Val Arg Thr Phe Val His Val Ser Thr Asp  
 115 120 125  
 Glu Val Tyr Gly Ser Leu Pro His Gly Ala Ala Ala Glu Ser Asp Pro  
 130 135 140  
 Leu Leu Pro Thr Ser Pro Tyr Ala Ala Ser Lys Ala Ala Ser Asp Leu  
 145 150 155 160  
 Met Ala Leu Ala His His Arg Thr His Gly Leu Asp Val Arg Val Thr  
 165 170 175  
 Arg Cys Ser Asn Asn Phe Gly Pro His Gln His Pro Glu Lys Leu Ile  
 180 185 190  
 Pro Arg Phe Leu Thr Ser Leu Leu Ser Gly Gly Thr Val Pro Leu Tyr  
 195 200 205  
 Gly Asp Gly Arg His Val Arg Asp Trp Leu His Val Asp Asp His Val  
 210 215 220  
 Arg Ala Val Glu Leu Val Arg Val Ser Gly Arg Pro Gly Glu Ile Tyr  
 225 230 235 240  
 Asn Ile Gly Gly Gly Thr Ser Leu Pro Asn Leu Glu Leu Thr His Arg  
 245 250 255  
 Leu Leu Ala Leu Cys Gly Ala Gly Pro Glu Arg Ile Val His Val Glu  
 260 265 270  
 Asn Arg Lys Gly His Asp Arg Arg Tyr Ala Val Asp His Ser Lys Ile  
 275 280 285  
 Thr Ala Glu Leu Gly Tyr Arg Pro Arg Thr Asp Phe Ala Thr Ala Leu  
 290 295 300  
 Ala Asp Thr Ala Lys Trp Tyr Glu Arg His Glu Asp Trp Trp Arg Pro  
 305 310 315 320  
 Leu Leu Ala Ala Thr  
 325

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 345 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Met	Thr	Met	Arg	Leu	Ala	Asp	Ile	Ala	Ala	Arg	Ala	Gly	Val	Ser	Lys
1				5					10					15	
Ala	Thr	Val	Ser	Arg	Val	Leu	Asn	Gly	Ala	Arg	Gly	Val	Ala	Thr	Gly
			20					25					30		
Thr	Arg	Asp	Ser	Val	Leu	Thr	Ala	Leu	Asp	Ala	Leu	Gly	Tyr	Glu	Gln
		35					40					45			
Pro	Ala	Arg	Gln	Arg	Arg	Ala	Glu	Leu	Val	Gly	Leu	Val	Thr	Pro	Glu
	50					55					60				
Leu	Asp	Asn	Pro	Phe	Phe	Pro	Ala	Leu	Ala	Gln	Val	Met	Gly	Gln	Ala
65					70					75					80
Leu	Thr	Arg	Gln	Gly	Tyr	Thr	Pro	Val	Leu	Ala	Thr	Gln	Thr	Pro	Gly
				85					90					95	
Gly	Ser	Thr	Glu	Asp	Glu	Leu	Thr	Glu	Met	Leu	Val	Asp	Arg	Gly	Val
			100					105					110		
Ser	Gly	Ile	Ile	Phe	Val	Ser	Gly	Leu	His	Ala	Asp	Thr	Thr	Ala	Glu
		115					120					125			
Thr	Gly	Arg	Tyr	Gly	Arg	Leu	His	Glu	Arg	Gln	Val	Pro	Phe	Val	Leu
	130					135					140				
Val	Asn	Gly	Phe	Ser	Pro	Arg	Ile	Glu	Ala	Pro	Phe	Val	Ser	Pro	Asp
145					150					155					160
Asp	Arg	Ala	Ala	Met	Arg	Leu	Ala	Val	Ala	His	Leu	Ala	Glu	Leu	Gly
				165					170					175	
His	Glu	Arg	Val	Gly	Leu	Ala	Val	Gly	Pro	Ala	Arg	Phe	Val	Pro	Val
			180					185					190		
Gln	Arg	Lys	Ile	Glu	Gly	Phe	Arg	Ala	Gly	Val	Arg	Glu	His	Leu	Gly
		195					200					205			
Val	Ser	Ala	Arg	Glu	Ser	Glu	Glu	Leu	Val	Gln	His	Ser	Leu	Phe	Ser
	210					215					220				
Leu	Glu	Gly	Gly	Gln	Ala	Ala	Ala	Ser	Ala	Leu	Ile	Asp	Leu	Gly	Cys
225				230						235					240
Thr	Ala	Val	Met	Cys	Ala	Ser	Asp	Met	Met	Ala	Leu	Gly	Ala	Val	Arg
			245						250					255	
Ala	Ala	Arg	Arg	Arg	Gly	Leu	Thr	Val	Pro	Gly	Asp	Ile	Ser	Val	Val
			260					265					270		
Gly	Phe	Asp	Asp	Ser	Pro	Leu	Met	Ala	Phe	Thr	Asp	Pro	Pro	Leu	Thr
		275					280					285			
Thr	Ile	Arg	Gln	Pro	Val	Lys	Ala	Met	Gly	Gln	Val	Ala	Val	Asp	Ala
	290					295					300				
Leu	Leu	Glu	Glu	Met	Ser	Gly	Thr	Pro	Pro	Pro	Arg	Thr	Glu	Phe	Val
305					310					315					320

Phe Met Pro Glu Leu Val Val Arg Gly Ser Thr Ala Ala Gly Pro Arg  
325 330 335

Gly Gly Arg Arg Pro Ala His Gly Arg  
340 345

(2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 393 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Met Arg Arg Gly Ile Ala Ala Thr Ala Leu Phe Ala Ala Val Ala Met  
1 5 10 15

Thr Ala Ser Ala Cys Gly Gly Gly Asp Asn Gly Gly Ser Gly Thr Asp  
20 25 30

Ala Gly Gly Thr Glu Leu Ser Gly Thr Val Thr Phe Trp Asp Thr Ser  
35 40 45

Asn Glu Ala Glu Lys Ala Thr Tyr Gln Ala Leu Ala Glu Gly Phe Glu  
50 55 60

Lys Glu His Pro Lys Val Asp Val Lys Tyr Val Asn Val Pro Phe Gly  
65 70 75 80

Glu Ala Asn Ala Lys Phe Lys Asn Ala Ala Gly Gly Asn Ser Gly Ala  
85 90 95

Pro Asp Val Met Arg Thr Glu Val Ala Trp Val Ala Asp Phe Ala Ser  
100 105 110

Ile Gly Tyr Leu Ala Pro Leu Asp Gly Thr Pro Ala Leu Asp Asp Gly  
115 120 125

Ser Asp His Leu Pro Gln Gly Gly Ser Thr Arg Tyr Glu Gly Lys Thr  
130 135 140

Tyr Ala Val Pro Gln Val Ile Asp Thr Leu Ala Leu Phe Tyr Asn Lys  
145 150 155 160

Glu Leu Leu Thr Lys Ala Gly Val Glu Val Pro Gly Ser Leu Ala Glu  
165 170 175

Leu Lys Thr Ala Ala Ala Glu Ile Thr Glu Lys Thr Gly Ala Ser Gly  
180 185 190

Leu Tyr Cys Gly Ala Thr Thr Arg Thr Trp Phe Leu Pro Tyr Leu Tyr  
195 200 205

Gly Glu Gly Gly Asp Leu Val Asp Glu Lys Asn Lys Thr Val Thr Val  
210 215 220

Asp Asp Glu Ala Gly Val Arg Ala Tyr Arg Val Ile Lys Asp Leu Val  
 225 230 235 240  
 Asp Ser Lys Ala Ala Ile Thr Asp Ala Ser Asp Gly Trp Asn Asn Met  
 245 250 255  
 Gln Asn Ala Phe Lys Ser Gly Lys Val Ala Met Met Val Asn Gly Pro  
 260 265 270  
 Trp Ala Ile Glu Asp Val Lys Ala Gly Ala Arg Phe Lys Asp Ala Gly  
 275 280 285  
 Asn Leu Gly Val Ala Pro Val Pro Ala Gly Ser Ala Gly Gln Gly Ser  
 290 295 300  
 Pro Gln Gly Gly Trp Asn Leu Ser Val Tyr Ala Gly Ser Lys Asn Leu  
 305 310 315 320  
 Asp Ala Ser Tyr Ala Phe Val Lys Tyr Met Ser Ser Ala Lys Val Gln  
 325 330 335  
 Gln Gln Thr Thr Glu Lys Leu Ser Leu Leu Pro Thr Arg Thr Ser Val  
 340 345 350  
 Tyr Glu Val Pro Ser Val Ala Asp Asn Glu Met Val Lys Phe Phe Lys  
 355 360 365  
 Pro Ala Val Asp Lys Ala Val Glu Arg Pro Trp Ile Ala Glu Gly Asn  
 370 375 380  
 Ala Leu Phe Glu Pro Ile Arg Leu Gln  
 385 390